Petaflops Special-Purpose Computer for Molecular Dynamics Simulations

Makoto Taiji

High-Performance Molecular Simulation Team

Computational & Experimental Systems Biology Group

Genomic Sciences Center, RIKEN

(Next-Generation Supercomputer R&D Center, RIKEN)

Acknowledgements

- For MDGRAPE-3 Project
  - Dr. Tetsu Narumi
  - Dr. Yousuke Ohno
  - Dr. Atsushi Suenaga
  - Dr. Noriaki Okimoto
  - Dr. Noriyuki Futatsugi
  - Ms. Ryoko Yanai

  - Ministry of Education, Culture, Sports, Science & Technology
  - Intel Corporation for early processor support
  - Japan SGI for system integration

Brief Introduction of RIKEN
(Institute of Physical and Chemical Research)

- Only research institute covers whole range of natural science and technology in Japan
- ~3,000 staffs
- Budget: ~700 million dollars/year
- 7 bioscience centers:
  - Genomic Sciences Center
  - SNP Research Center
  - Plant Science Center
  - Center for Allergy and Immunology
  - Brain Science Institute
  - Center for developmental biology
  - BioResource Center
- Next-Generation Supercomputer (10PFLOPS at FY2011)
- Genomic Science Center:
  - The most important national center of genome/post-genome research
  - National projects
    - Protein 3000 Project
    - ENU Mouse mutagenesis
    - Genome Network Project

What is GRAPE?

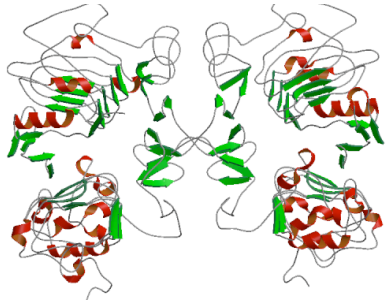
- GRAvity PipE
- Special-purpose accelerator for classical particle simulations
  - Astrophysical *N*-body simulations
  - Molecular Dynamics Simulations
- MDGRAPE-3 : Petaflops GRAPE for Molecular Dynamics simulations

J. Makino & M. Taiji, *Scientific Simulations with Special-Purpose Computers,* John Wiley & Sons, 1997.

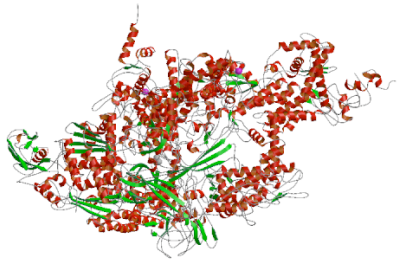
MDGRAPE-3 (aka Protein Explorer)

- Petaflops special-purpose computer for molecular dynamics simulations
- Started at April 2002, Finished at June 2006
- Part of Protein 3000 project – a project to determine 3,000 protein structures

RIKEN NMR Park



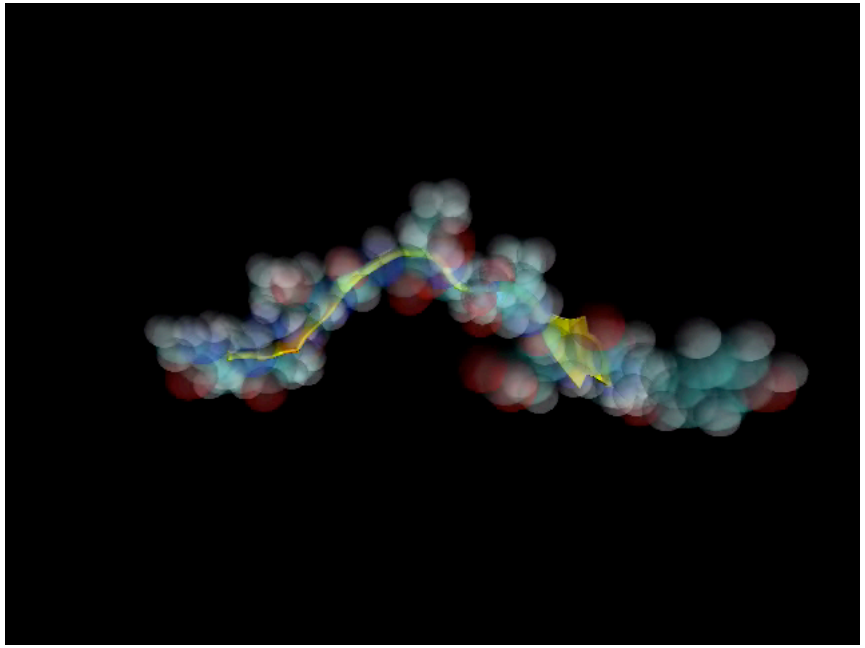
EGFR



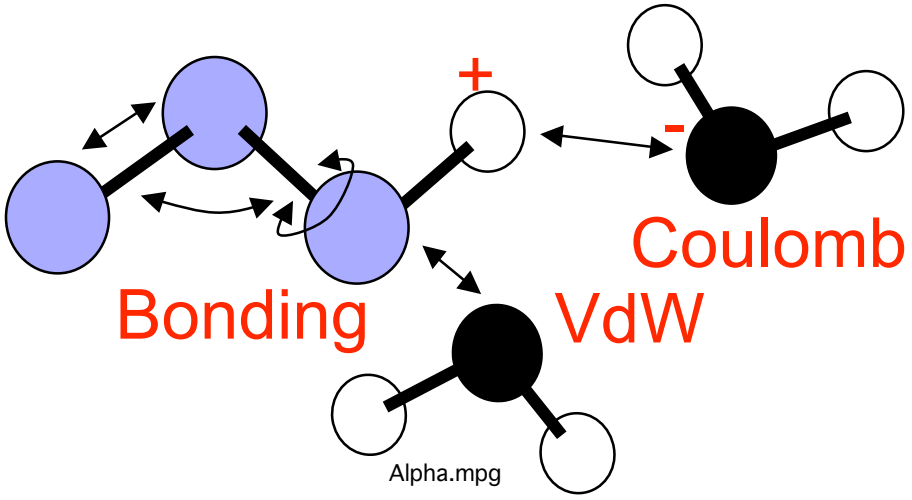
TT RNA Polymerase

M. Taiji et al, Proc. Supercomputing 2003, on CDROM.
M. Taiji, Proc. Hot Chips 16, on CDROM (2004).

Molecular Dynamics Simulations



Folding of Chignolin,
10-residue β-hairpin
design peptide
(by Dr. A. Suenaga)

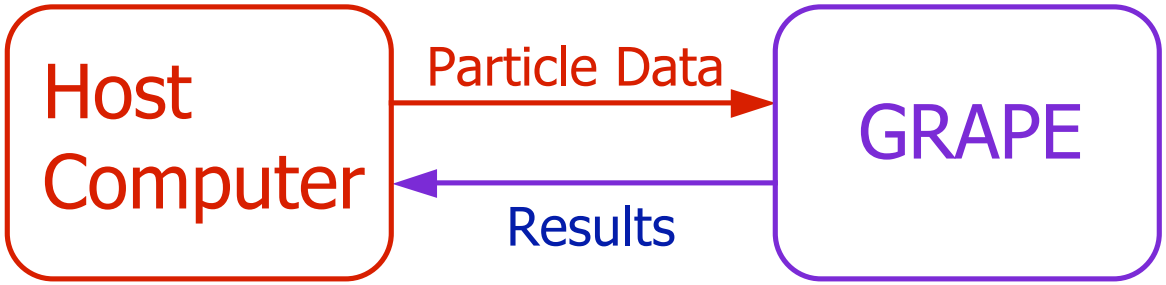


Force calculation dominates computational time

Require large computational power

How GRAPE works

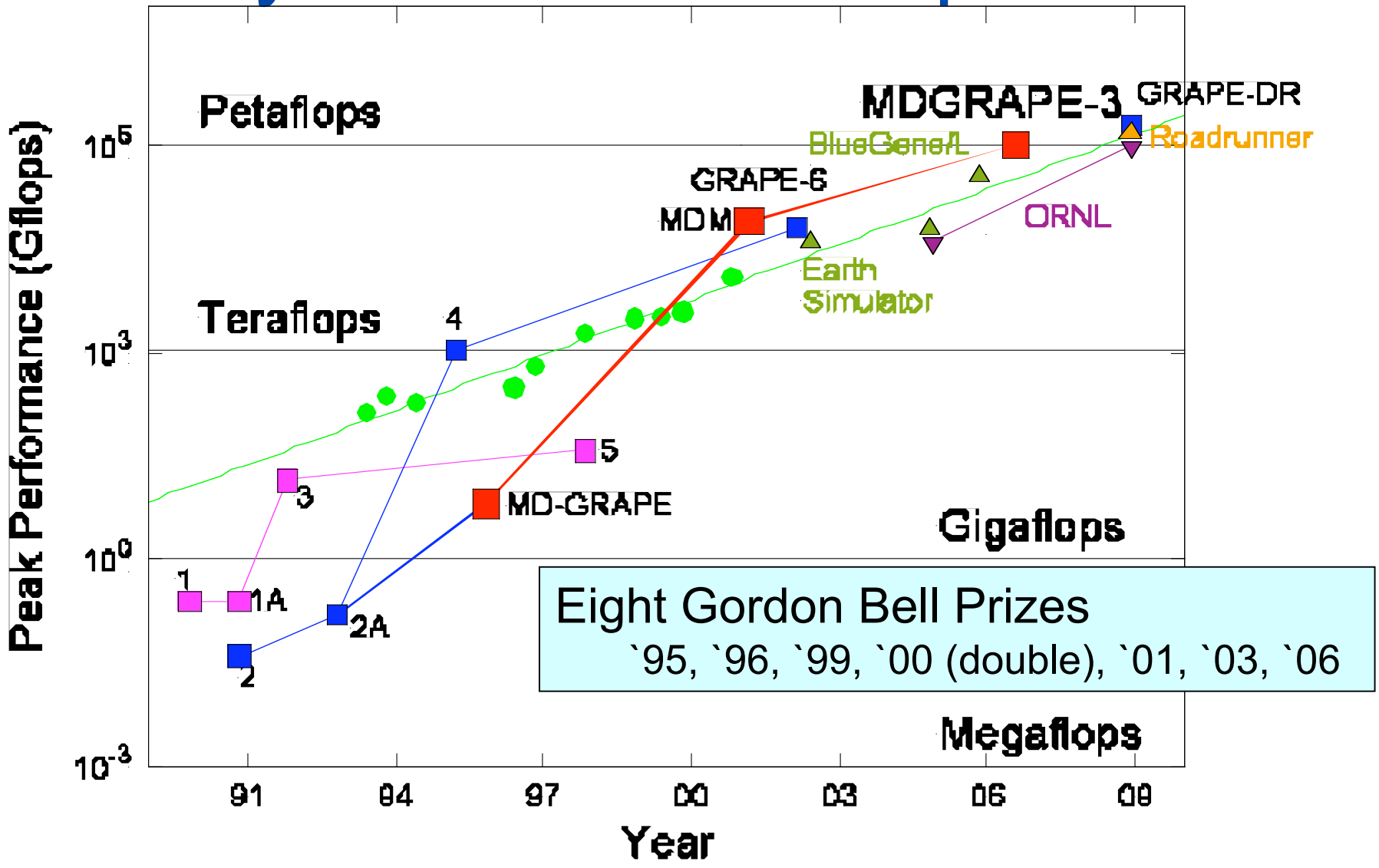
- Accelerator to calculate forces



Most of Calculation → GRAPE
Others → Host computer

- Communication = O (N) << Calculation = O (N^2)
- Easy to build, Easy to use
- Cost Effective

History of GRAPE computers



Why we build special-purpose computers?

Bottleneck of high-performance computing:

- Parallelization limit / Memory bandwidth
- Power Consumption = Heat Dissipation

These problems will become more serious in future.

Special-purpose approach:

- can solve parallelization limit for some applications
- relax power consumption
- ~100 times better cost-performance

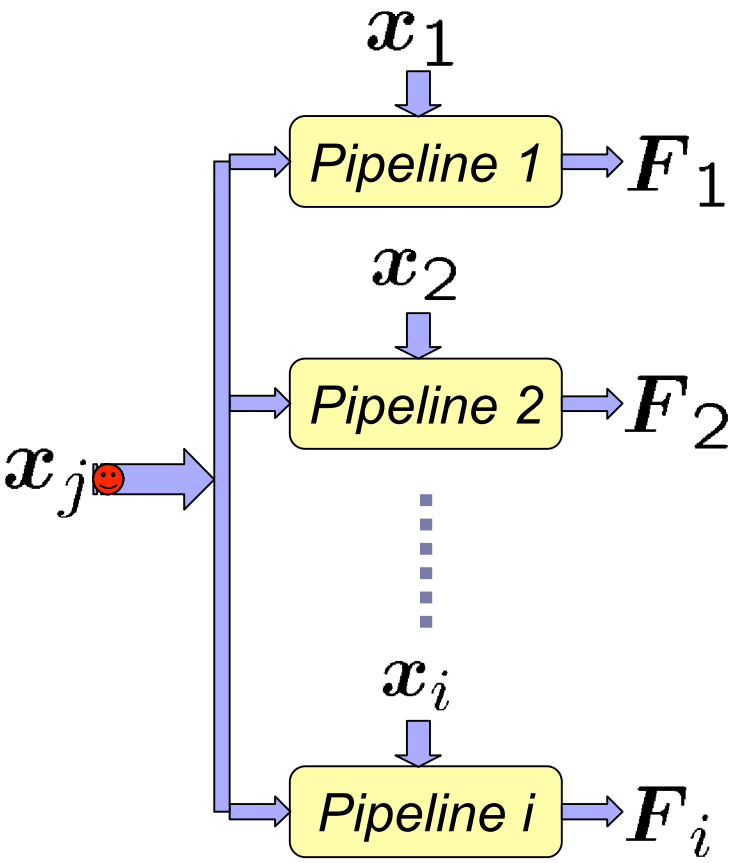
Broadcast Parallelization

- Molecular Dynamics Case
- Two-body forces

$$\boldsymbol{F}_i = \sum_j \boldsymbol{f}(x_i, x_j)$$

- For parallel calculation of $\boldsymbol{F}_i$, we can use the same $\boldsymbol{x}_j$

- Broadcast Parallelization
  - relax Bandwidth Problem



Highly-Parallel Operations in Molecular Dynamics Processors

- For special-purpose computers
  - Broadcast Memory Architecture
  - Efficient : 720 operations/cycle/chip
    in MDGRAPE-3 chip
  - possible to increase according to Moore's law
- In case of MD:

MDGRAPE	600 nm	1 pipeline	1Gflops
MDGRAPE-2	250 nm	4 pipelines	16Gflops
MDGRAPE-3	130 nm	20 pipelines	180Gflops

Power Efficiency of Special-Purpose Computers

- If we compare at the same technology
  - Pentium 4 (0.13 μm, 3GHz, FSB800) … 14W/Gflops
  - MDGRAPE-3 chip (0.13μm) … 0.1W/Gflops
- Why ?
  - Highly-parallel at low frequency

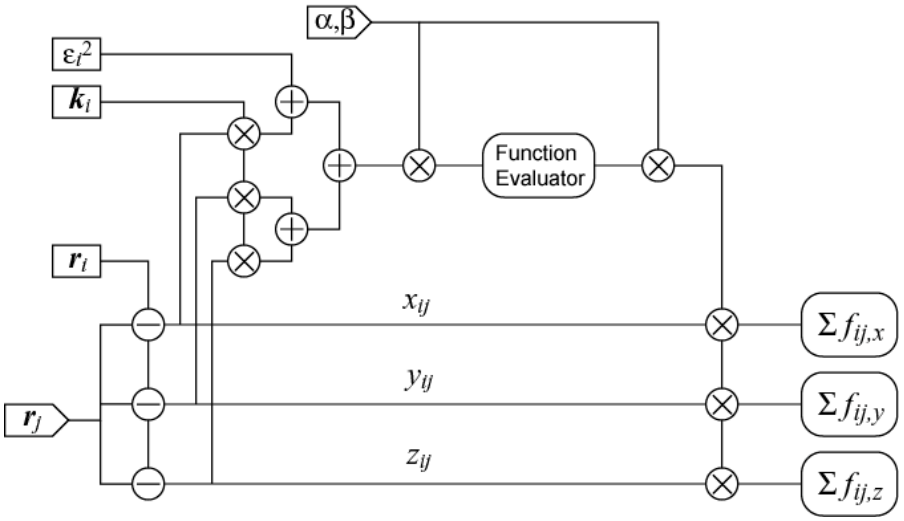
    MDGRAPE-3: 250MHz, 720-equivalent operations

    for example, single-precision multiplier has 3 pipeline stages
  - Tuning accuracy

    Most of calculations are done in single precision
  - Slow I/O

    84-bit wide input and output port at 125 MHz (GTL)

Force Pipeline



- Calculate two-body central forces

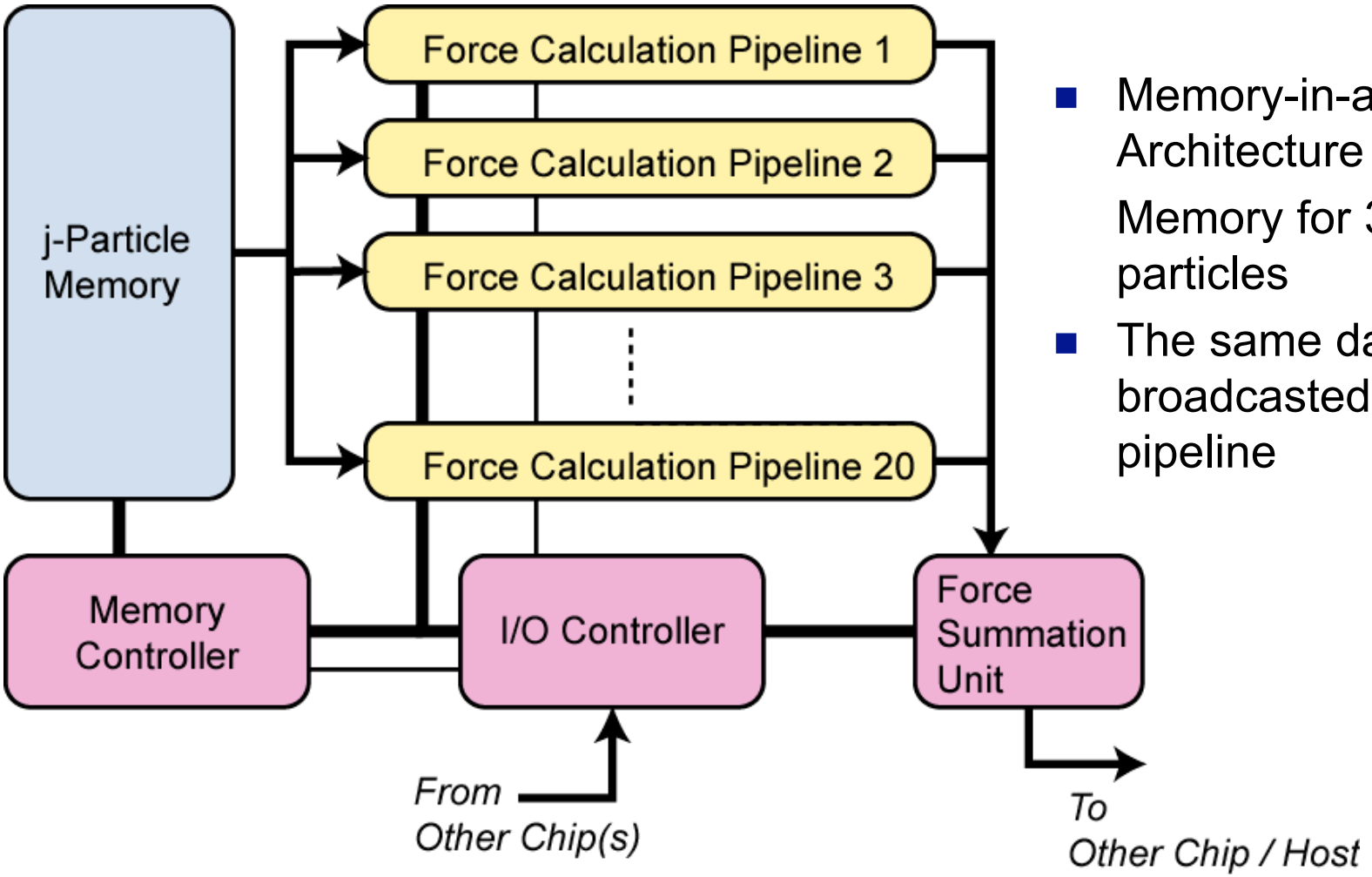
$$\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$$

$$r_{ij}^2 = r_{ij}^2 + \varepsilon^2$$

$$\mathbf{F}_i = \sum_j r_{ij} \, g(r_{ij}^2)$$

- 8 multipliers, 9 adders, and 1 function evaluator = 33 equivalent operations for Coulomb force calculation
  A. H. Karp, *Scientific Programming*, **1**, pp133–141 (1992)
- Function Evaluator: approximate arbitrary functions by segmented fourth-order polynomials
- Multipliers : floating-point, single precision
- Adders: floating-point, single precision / fixed-point 40 or 80 bit

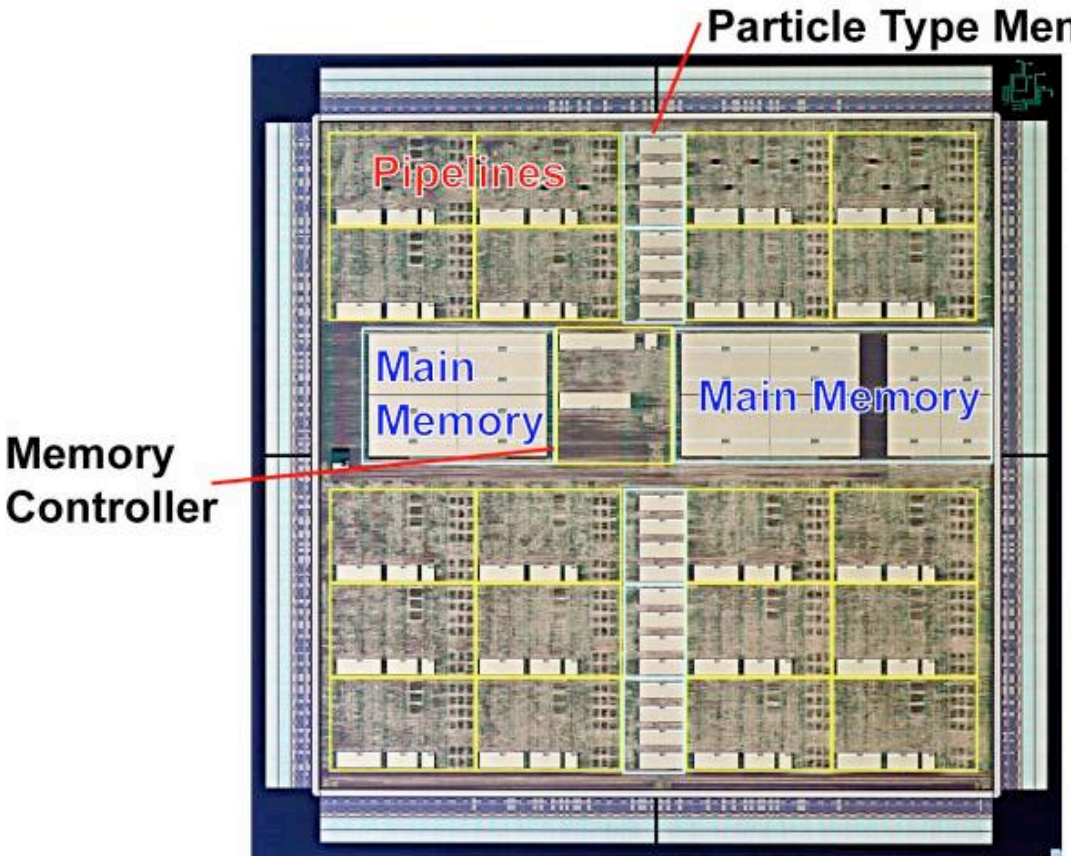
Block Diagram of MDGRAPE-3 chip



- Memory-in-a-chip Architecture

  Memory for 32,768 particles
- The same data is broadcasted to each pipeline

MDGRAPE-3 chip

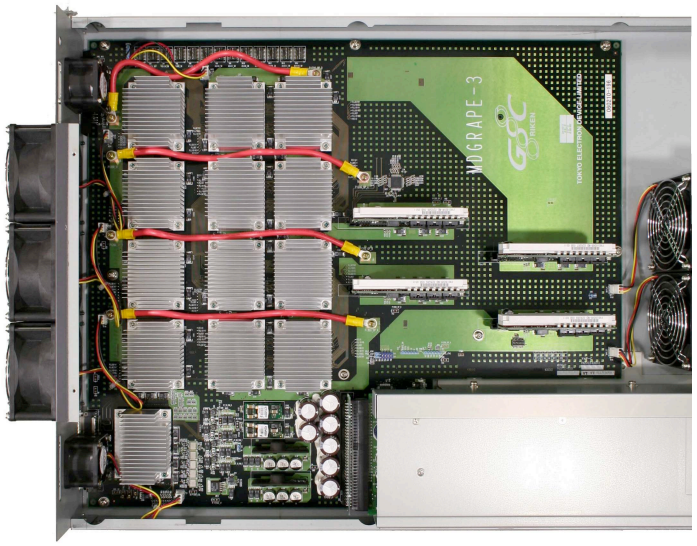


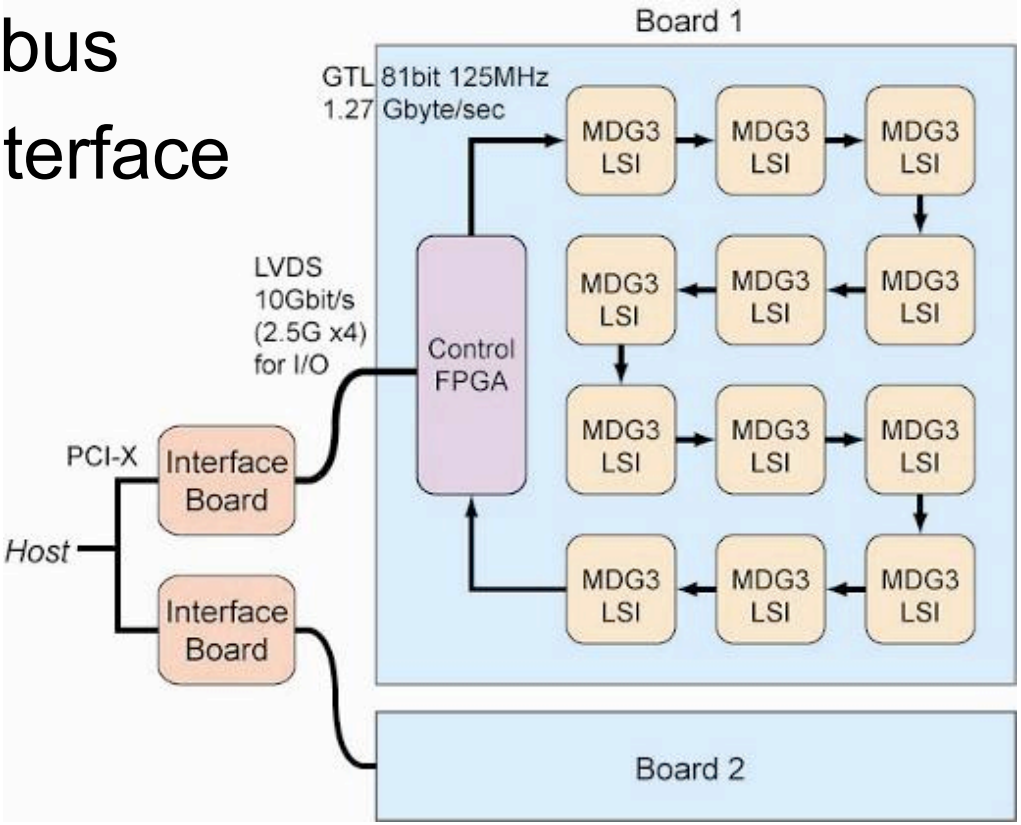
216 GFLOPS@300MHz
180 GFLOPS@250MHz
17W at 300 MHz

Hitachi HDL4N
130 nm
Vcore=+1.2V
15.7 mm X 15.7 mm
6.1 M random gates
+ 9 Mbit memory
1444 pin FCBGA

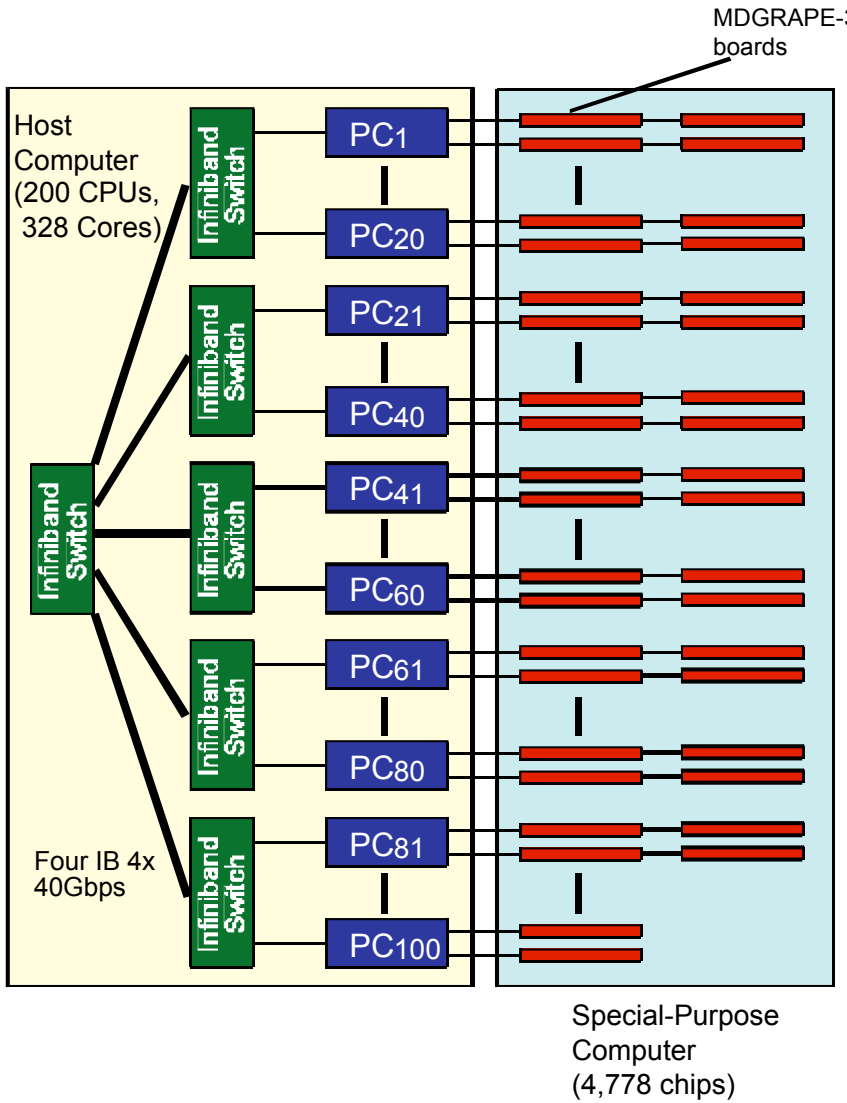
MDGRAPE-3 Board

- 12 Chips/Board
- 2 boards/2U subrack = 5 Tflops
- Connected to PCI-X bus
  via LVDS 10Gbit/s interface





MDGRAPE-3 system

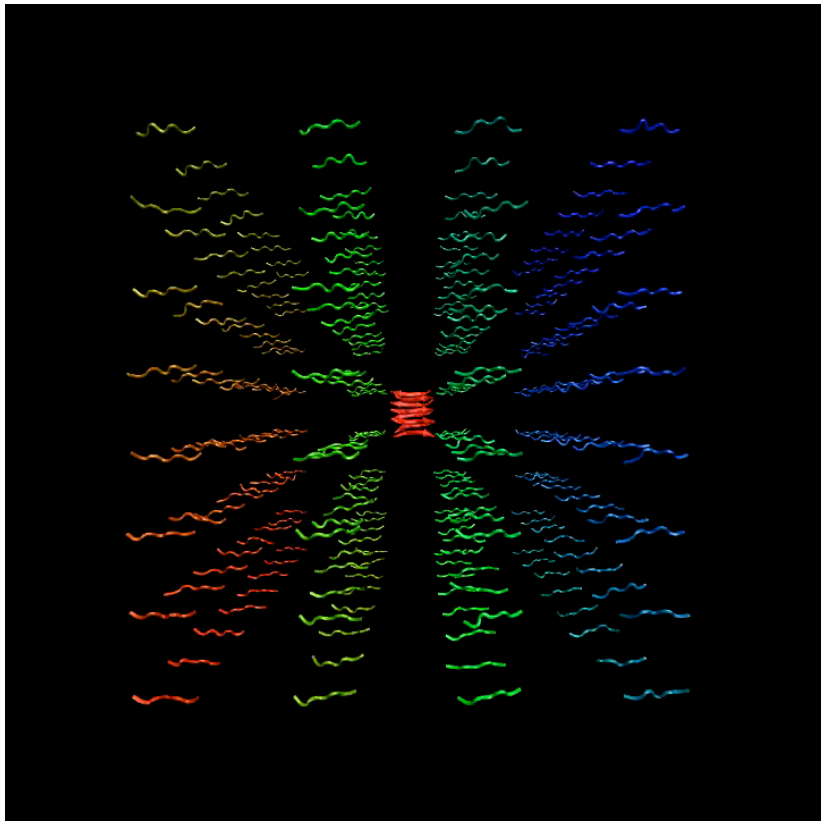


- 4,778 dedicated LSI “MDGRAPE-3 chip”
  - 300MHz(216Gflops) 3,890
  - 250MHz(180Gflops) 888
- Nominal Peak Performance: 1 Petaflops
- Total 400 boards with 12(some 11) MDGRAPE-3 chips
- Host : Intel Xeon Cluster, 370 cores
  - Dual-core Xeon 5150(Woodcrest 2.66GHz) 2way server x 85 Nodes provided by Intel Corporation
  - Xeon 3.2DGHz 2way server x 15 Nodes
  - System Integration: Japan SGI
- Power Consumption : 200kW
- Size : 22 standard 19inch racks
- Cost : 8.6 M$ (including Labor)

MDGRAPE-3 system

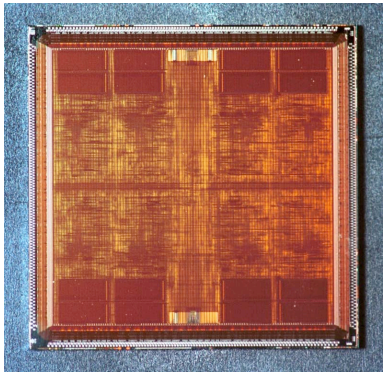
Sustained Performance of Parallel System

- Gordon Bell 2006 Honorable Mention, Peak Performance
- Amyloid forming process of Yeast Sup 35 peptides
- Systems with 17 million atoms
- Cutoff simulations
  (Rcut = 45 Å)
- Nominal peak : 860 Tflops
- Running speed : 370 Tflops
- Sustained performance: 185 Tflops
- Efficiency ~ 45 %



Applications suitable for broadcast memory architecture

- Multiple calculations using the same data
  - Molecular dynamics / Astrophysical *N*-body simulations
  - Dynamic programming for genome sequence analysis
  - Boundary value problems
  - Calculation of dense matrices(incl. Linpack)
- SIMD (vector) processor with broadcast memory architecture
  - MACE (MAtrix Computing Engine)
    for dense matirix calculation
    3.5Gflops/chip, double precision, 180nm
  - GRAPE-DR Project (2004-2009)

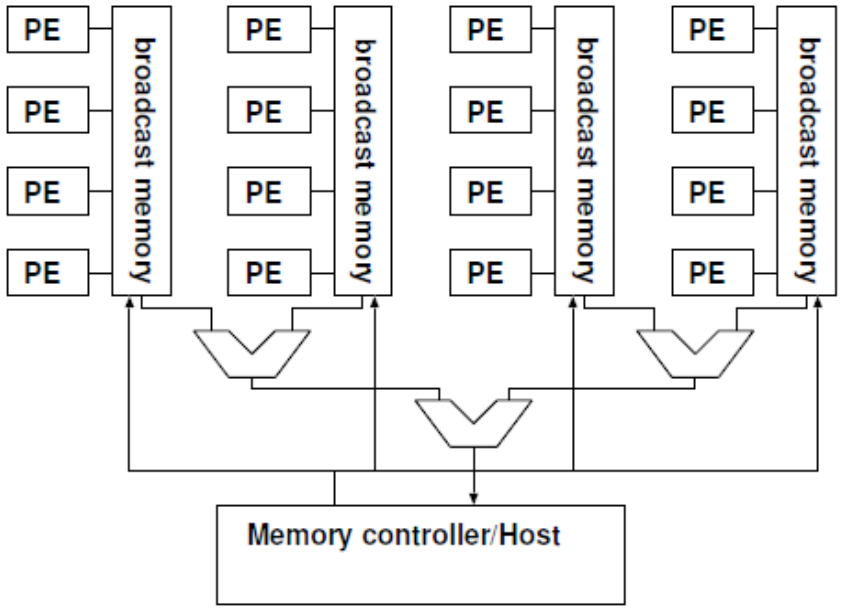


GRAPE-DR Project

- Greatly Reduced Array of Processor Elements with Data Reduction
- SIMD accelerator with broadcast memory architecture
- Full system: FY2008
- 0.5 TFLOPS / chip (single), 0.25 TFLOPS (double)
- 2 PFLOPS / system
- Prof. Kei Hiraki (U. Tokyo)
  Prof. J. Makino (National Astronomical Observatory)
  Dr. T. Ebisuzaki (RIKEN)

SING (SING is not GRAPE) chip

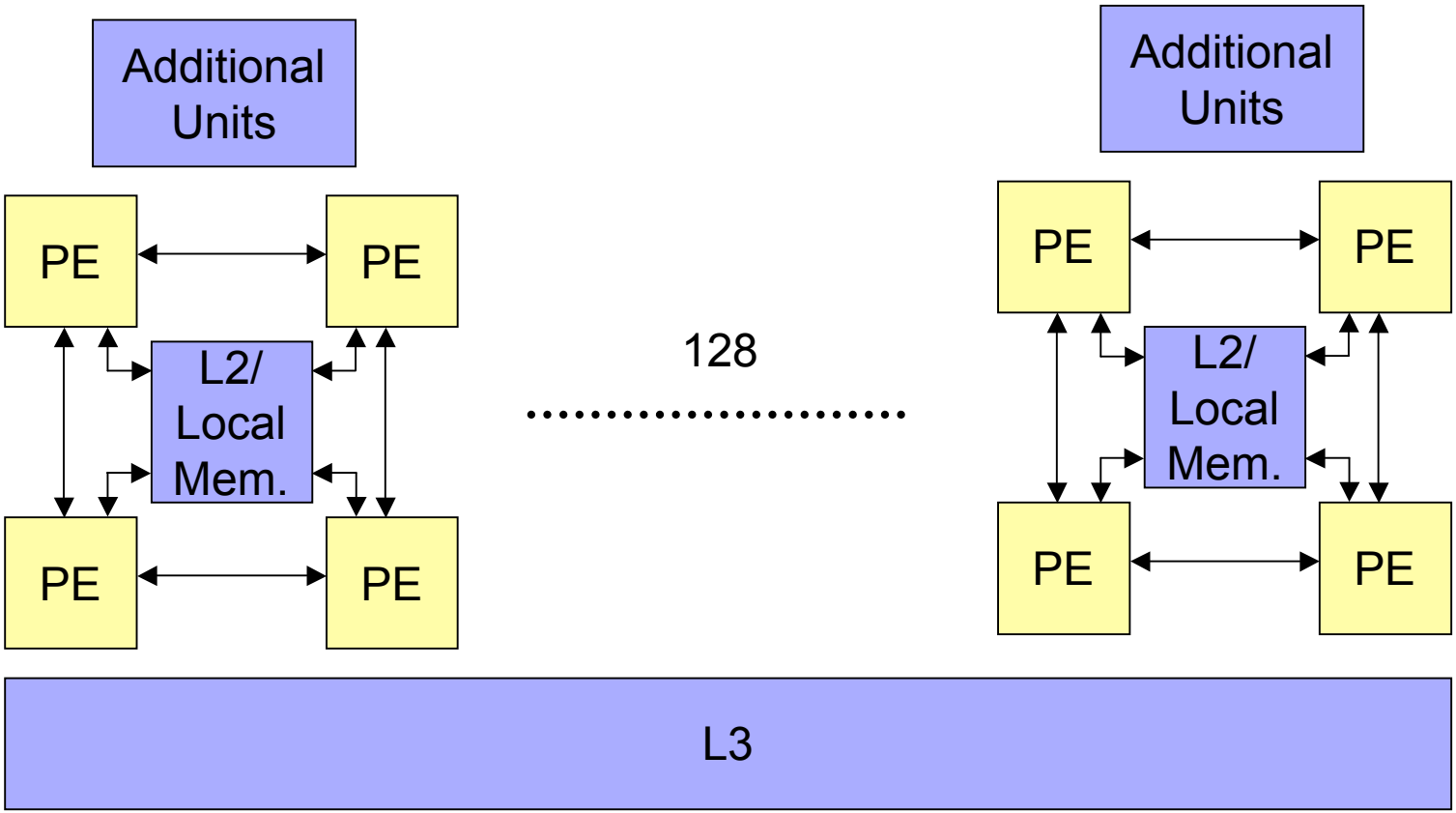
- 512 Processor Elements, 500 MHz
- PE:
  - FP Mul/Add
  - Integer ALU
  - 32-word Register File
  - 256-word memory
- 0.5 TFLOPS, ~0.1W/GFLOPS(SP)
- 0.25TFLOPS, ~0.2W/GFLOPS(DP)



J. Makino et al.,
http://www.ccs.tsukuba.ac.jp/workshop/sympo-060404/pdf/3-7.pdf *(in japanese)*

MDGRAPE-4: combination of dedicated and general-purpose units

- SIMD Accelerator with broadcast memory architecture
  Problem: too many parallelism
  ~500/chip, 5M/system - Works with SIMD?
- What is good with dedicated pipelines
- Force calculation ~ 30 operations done by pipelined operations
  Systolic computing
- Can decrease parallelism
- VLIW-like (SIMD) processor with chained operation can mimic pipelined operations
- Allows to embed more dedicated units
  which can not be fully utilized by SIMD operations



Each PE: Simple in-order processor with L1$

Additional Units can be:

Lookup table
(for polynomial interpolations or VdW coefficients),
1/x, Function evaluator etc.

Target: ~ 0.1W/GFLOPS (DP)

Summary

- MDGRAPE-3 achieved PetaFLOPS nominal peak for 200 kW
- Dedicated parallel pipelines at modest speed of ~250 MHz results high performance/power
- Generalized GRAPE approaches are being developed